STOCHASTIC PROCESSES AND STATISTICS

By J. L. Doob*

Department of Mathematics, Columbia University

Communicated May 5, 1934

A stochastic process is defined by Khintchine[1] to be a one parameter set of chance variables: $\mathbf{x}(t)$, $-\infty < t < \infty$. It is supposed that if $t_1, \ldots, t_n$ is any finite set of values of t, and $a_j < x < b_j$, $j = 1, \ldots, n$ any set of intervals, the probability that

$$a_j < \mathbf{x}(t_j) < b_j, \; j = 1, \ldots, n, \tag{1}$$

is defined. If the probability that (1) is true is independent of translations of the t-axis, the process is called stationary.

Let Ω be any space in which a probability measure[2] is defined. Let $\{T_t\}$, $-\infty < t < \infty$, be a set of one to one transformations of Ω into itself, such that $T_{t_1+t_2} = T_{t_1}T_{t_2}$. If ω is an element of Ω, the set of elements $\{T_t\omega\}$, $-\infty < t < \infty$ is called a path curve. These path curves have been studied, and many results, such as the ergodic theorem, obtained. It was shown by Khintchine[1] that there is a formal analogy between this study and that of stochastic processes. It will be shown in this paper that the two theories are abstractly identical.

An example of a stochastic process can be obtained as follows. Let Ω, T_t be defined as in the preceding paragraph, and let $\varphi(\omega)$ be a measurable function on Ω. A chance variable is, by definition, a measurable function on a space on which a probability measure is defined. We define $\mathbf{x}(t)$ as the function $\varphi(T_t\omega)$. The probability that (1) is true is then the measure of the set of elements ω such that

$$a_j \leq \varphi(T_{t_j}\omega) < b_j, \quad j = 1, \ldots, n. \tag{2}$$

The following theorem shows that every stochastic process can be obtained in this way. If the stochastic process is stationary, the transformations $\{T_t\}$ are measure preserving, and conversely.

THEOREM 1. *A stochastic process can always be considered as a set of measurable functionals $\varphi_t(\omega)$, $-\infty < t < \infty$ on the function space Ω of functions $x(t) = \omega$ defined for $-\infty < t < \infty$, on which a probability measure is defined. If T is the transformation of Ω taking $x(t)$ into $x(t + \tau)$, $\varphi_t(\omega) = \varphi_0(T_t\omega)$.*

Let $\{\mathbf{x}(t)\}$ be the chance variables of the given stochastic process. There corresponds to every value of t a space Ω_t of elements ω_t on which a probability measure is defined, and $\mathbf{x}(t)$ is a measurable function, $\varphi_t(\omega_t)$, on Ω_t. Transform Ω_t into the x-axis by the transformation S_t which takes the set of points ω_t at which $\varphi_t(\omega_t) = a$ into the point $x = a$ for each value of a. Then $\varphi_t(\omega_t) = x$ if $x = S_t\omega_t$, and the probability measure on

Ω_t induces one on the x-axis. The chance variable $\mathbf{x}(t)$ becomes the chance variable $\mathbf{x}^*(t) = x$. We can thus suppose that $\mathbf{x}(t)$ is a chance variable corresponding to a distribution on the x-axis, and that $\varphi_t(\omega_t) = \varphi_t(x) = x(t)$, i.e., that $\varphi_t(\omega_t)$ is the function x—where measure is defined in a way depending on t. By hypothesis, probability is assigned to events determined by conditions of the form

$$a_j \leq x(t_j) < b_j, \quad j = 1, \ldots, n, \tag{1'}$$

i.e., a probability measure is defined on the function space Ω, the measure being determined by its values on sets determined by conditions of the form (1').[3] The chance variable $\mathbf{x}(t_0)$ can be considered as the function defined on Ω which takes on the value $x(t_0)$ at the element $x(t)$ of Ω. This functional is a measurable function on Ω. The first part of the theorem is thus proved. The second part is obvious.

Since $x(t + \tau) = T_\tau[x(t)]$, Theorem 1 shows that if the process is stationary, and if simple continuity conditions are satisfied,[4] the ergodic theorem of Birkhoff and similar theorems can be applied. For instance, by the ergodic theorem, if $\mathbf{x}(0)$ (considered as a function on Ω) is integrable over Ω (i.e., if the expectation of the chance variable $\mathbf{x}(0)$ exists) there is a chance variable $\mathbf{x}$ such that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{x}(t)dt = \mathbf{x} \tag{3}$$

with probability 1.[5]

If we suppose that the chance variable $\mathbf{x}(t)$ of a stochastic process is independent of t for $n \leq t < n + 1$ for every integer n, the specialization of the above has important applications. The process then becomes essentially a sequence $\ldots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \ldots$ of chance variables, or, as in Theorem 1, the stochastic process becomes a set of measurable functions defined on the space Ω_s of all sequences $(\ldots, x_{-1}, x_0, x_1, \ldots)$ on which a probability measure is defined.[6] Here (3) becomes

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j = \mathbf{x}^{7} \tag{3'}$$

true with probability 1, if the process is stationary, which means here that the probability measure on Ω_s is to be invariant under the transformation T taking $(\ldots, x_{-1}, x_0, x_1, \ldots)$ into $(\ldots, x_0, x_1, x_2, \ldots)$, and if the expectation of $\mathbf{x}_1$ exists.

The special case in which the chance variables $\mathbf{x}_0, \mathbf{x}_{\pm 1}, \ldots$ are independent and all have the same distributions is of importance. This process is stationary, so (3') holds. It can be proved that in this case the transformation T is metrically transitive. This means that $\mathbf{x}$ is a constant[8]—in fact that $\mathbf{x}$ is the expectation of $\mathbf{x}_1$. Conversely the following theorem can be proved.

THEOREM 2. *Let* $\mathbf{x}_1, \mathbf{x}_2, \ldots$ *be a sequence of independent chance variables with the same distributions. If there is a sequence of constants* $c_1, c_2, \ldots$ *such that the probability that*

$$\limsup_{n\to\infty} \left| \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j - c_n \right| < \infty \tag{4}$$

is greater than zero, then the expectation E *of* $\mathbf{x}_j$ *exists, and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j = E \tag{3''}$$

with probability 1.

This theorem depends on the fact that if $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is a sequence of independent chance variables with the same distributions, a necessary and sufficient condition that their expectations exist is that the probability that $\limsup_{n\to\infty} |\mathbf{x}_n/n| < \infty$ is greater than 0.

This work can be considered as a generalization of the following fact. If $f(x)$ is the probability density[9] of a chance variable $\mathbf{x}$, $\prod_{j=1}^{n} f(x_j)$ is the density for the distribution of the results of n independent trials. What has been done above can be interpreted as letting n become infinite, obtaining a probability set-up which is suitable for any finite number of trials.

Using the results described above, a rigorous proof can be obtained of the validity of the method of maximum likelihood of R. A. Fisher, which has supplanted the use of Bayes' Theorem.[10] Let $f(x, p)$ for each value of p in a neighborhood of a value p_0 be a probability density. Let $\mathbf{x}$ be a chance variable whose distribution has density $f(x, p_0)$. The problem is to estimate p_0 by means of samples of values of x. The method of Fisher is to choose the value $p_n(x_1, \ldots, x_n)$ of p which maximizes $\prod_{j=1}^{n} f(x_0, p)$ for fixed $x_1, \ldots, x_n$. Then it can be shown using the above results, that under suitable restrictions on the continuity of $f(x, p)$ in p, $p_n(x_1, \ldots, x_n)$ approaches p_0 with probability 1, and that for large n the distribution of $p_n(x_1, \ldots, x_n)$ is nearly normal, with mean p_0 and variance $1/n\sigma^2$, where

$$\sigma^2 = -\int_{-\infty}^{\infty} f(x, p_0)\, \frac{\partial^2}{\partial p^2} \left[(\log f(x, p_0) \right] dx.$$

[1] *Mathemat. Ann.*, **109**, 604–615 (1934); this paper should be compared with a paper by the same author in these PROCEEDINGS, **19**, 567–573 (1933).

[2] We shall mean by this that there is a non-negative completely additive set function, defined on a collection of sets of elements of the space (called measurable sets). If $A_1, A_2, \ldots$ are measurable sets, we suppose that their complements and their sum are measurable. We suppose that the set of all elements in the space has measure 1. Measurable functions and Lebesgue integration can be defined in the usual way.

[3] This way of defining measure in function space was discussed by Kolmogoroff, *Ergebnisse Mathematik*, **2**, No. 3, Grundbegriffe der Wahrscheinlichkeitsrechnung, § 4.

[4] It is sufficient that if E is any set in Ω determined by conditions of the form (1′), and if E is transformed into E_t by T_t, the measure of $E \cdot E_t$ should be continuous in t at $t = 0$.

[5] For a simple proof of the ergodic theorem, following the lines of the first proof, given by Birkhoff, cf. A. Khintchine, *Mathemat. Ann.*, **107**, 485–488 (1933).

[6] This situation was discussed by Khintchine, *Zeit. Angewandte Mathemat. Mechanik*, **13**, 101–103 (1933), who treated the particular case of chance variables taking on only the values 1 or 0. The general case was discussed by E. Hopf, *Journal of Mathematics and Physics*, M. I. T., **13**, 51–102 (1934), who obtained (3′) but not Theorem 2.

[7] Kolmogoroff, loc. cit.,[3] p. 59, announced this result in the special case of independent chance variables, and announced also Theorem 2, under the assumption that the probability is 1 that the upper limit in (4) is 0.

[8] Loc. cit.,[5] p. 488.

[9] If $f(x)$ is defined for $-\infty < x < \infty$ except possibly for a set of points of Lebesgue measure 0, is Lebesgue measurable, not negative, and integrable over $-\infty < x < \infty$, $\int_{-\infty}^{\infty} f(x)dx = 1$, $f(x)$ will be called a probability density.

[10] This method was discussed (unrigorously) by Fisher in the *Phil. Trans. Roy. Soc. London*, Series A, **222** (1921). The treatment of H. Hotelling, *Trans. Amer. Math. Soc.*, **32**, 847–859 (1930), holds only in certain special cases.

* NATIONAL RESEARCH FELLOW.

REMARKS ON THE POSSIBLE FAILURE OF ENERGY CONSERVATION

BY RICHARD C. TOLMAN

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated April 17, 1934

1. *Possibility of Failure in the Case of Elementary Processes.*—The possible failure of the principle of the conservation of energy, in the case of the continuous β-ray spectrum accompanying radioactive decompositions, and perhaps also in the case of processes occurring in the interior of stars, has several times been suggested by Bohr.[1] From a theoretical point of view such a failure might be due to a breakdown in the applicability of ordinary mechanical notions under circumstances where the electron would have to be regarded as localizable within regions small compared with its classical dimensions.

In the case of the stars there is at present no definite observational evidence which would lead us to abandon the principle of the conservation of energy, beyond the removal of limitations on our attempts to explain the continued luminosity of those objects and to account, in general, for the existence of a supply of available energy in the universe.